IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 March 2023, accepted 17 April 2023, date of publication 19 April 2023, date of current version 2 May 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3268551*

RESEARCH ARTICLE

# User Biometric Identification Methodology via EEG-Based Motor Imagery Signals

**SUJIN BAK[1] AND JICHAI JEONG[2], (Senior Member, IEEE)**
[1]Advanced Institute of Convergence Technology, Yeongtong-gu, Suwon-si, Gyeonggi-do 16229, Republic of Korea
[2]Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea

Corresponding author: Jichai Jeong (jcj@korea.ac.kr)

This work was supported in part by the Advanced Institutes of Convergence Technology under Grant AICT-2023-0003, and in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korea Government (MSIT) (Development of BCI-Based Brain and Cognitive Computing Technology for Recognizing User's Intentions Using Deep Learning) under Grant 2017-0-00451.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Korea University Institutional Review Board under Application No. 1040548-KUIRB-16-159-A-2, and performed in line with the declaration of Helsinki.

**ABSTRACT** Human brain activities—electroencephalogram (EEG) signals—are likely to provide a secure biometric approach for user identification because they are more sensitive, secretive, and difficult to replicate. Many studies have recently focused on identifying and quantifying important frequency patterns in motor imagery (MI), recorded through EEG. However, there is still a lack of an optimal methodology for recognizing users with EEG-based MI. Therefore, we aimed to propose an EEG-MI methodology that utilizes optimized feature extraction methods and classifiers to improve user-aware accuracy. To accomplish this goal, we extracted four features related to MI and compared the accuracies for recognizing users using a support vector machine (SVM) and Gaussian Naïve Bayes (GNB). We then used the half-total error rate (HTER) to determine whether the results were reliable due to an imbalance problem caused by the differences in the data sizes. Thus, we used a common spatial pattern (CSP) to achieve the highest user identification accuracies of 98.97% and 97.47% using SVM and GNB, respectively. All user recognition accuracies are guaranteed by the HTERs, which are below 0.5. However, CSP has the disadvantage of decreasing accuracy on a small dataset scale. Therefore, we proposed and tested a statistical methodology for estimating a minimum dataset scale to ensure CSP performance. We confirm that the used dataset adequately guarantees CSP performance. This study makes a great contribution to the field of information security by presenting an EEG-MI methodology that improves the identification accuracy in human biometrics based on EEG-MI signals.

**INDEX TERMS** Biometric, electroencephalography (EEG), motor imagery (MI), support vector machine (SVM), user identification methodology, Gaussian Naïve Bayes (GNB).

## I. INTRODUCTION

User identification methods have rapidly advanced, and they are essential in many information-security fields. This method frequently requires the user's credentials to identify the user, while communicating between humans and computers. Recently, biometrics has gradually emerged to enhance the security of the user's identification method. They are designed to extract the physiological features obtained from a user using signal processing, machine learning, and pattern recognition techniques, and these features are then compared to the user's profile/template stored in a database. These biometric characteristics include the face, iris, voice, gait, finger or footprint, and signature technologies [1]. However, there is still a risk of forgery and theft of the biometric characteristics from users. For example, physical characteristics such as

The associate editor coordinating the review of this manuscript and approving it for publication was Tony Thomas.

photos of faces or voice recordings can be digitally captured, copied, or forged by an attacker [2]. A study has demonstrated the possibility of generating master prints that can be used by an adversary to launch a dictionary attack against a fingerprint recognition system [3]. In addition, biometric systems are vulnerable to external attacks, including attacks by individuals wearing contact lenses with printed artificial textures [4] and fake iris images [5].

To prepare for these security attacks, new biometric identification methodologies using EEG have been developed and proposed [6], [7], [8]. There are various reasons why EEG is a reasonable indicator for biometrics. A mental passphrase is difficult to steal because brain activity is unique for each individual and stable over time [9], [10], [11]. Some studies have shown that a user's mental passphrase is well-defended against attackers [12], [13], [14]. Moreover, as hardware technologies have advanced gradually, simplified EEG devices have been developed, such as Dry EEG [15], NeuroSky [16], and Emotiv EEG headsets [17]. All of which, have portability and communication capabilities. Furthermore, because of the increase in miniaturized and mass-produced EEG products, user identification methods that use EEG have become more common [18], [19], [20], [21].

Regarding EEG, biometrics have traditionally been used with a variety of strategies that include specific brain regions associated with human behavior. These strategies include vision, speech, auditory signals, and motor imagery (MI). One of the most commonly used tasks for recognizing users is visual recognition [22] because it has the advantage of providing quick responses. Nevertheless, there is always a risk of forgery and loss of media, such as face photography. Visual recognition is also not suitable for individuals who have lost their sight naturally. Thus, recent studies commonly suggest that MI is a more appropriate strategy [23], [24], [25], [26]. Rig Das et al. proposed EEG-based biometric identification, using a convolutional neural network (CNN) known as one of deep learning method omitting feature extraction procedures [24]. MI strategies can elicit brain activity while imagining moving the body, which not only generates the most powerful detectable patterns but also indicates that MI is significantly more accurate than visual imagery [27].

In the processing of brain signals caused by the MI strategy, feature extraction methods are regarded as one of the very important steps. Common spatial pattern (CSP) and event-related (de)synchronization (ERD/S) are the representative MI features extraction methods. Indeed, CSP shows that the frequency response components occurring in the motor cortex regions represent different spatiotemporal patterns among subjects [28]. Each subject possesses different frequency response components because of individual brain positions, signal patterns, and frequency bands [29]. These characteristics facilitate user identification. Moreover, MI is accompanied by both ERD and ERS. Indeed, ERD shows the suppression of electrode activity amplitudes, while ERS demonstrates the enhancement of electrode activity amplitudes over primary sensorimotor areas [30]. Both, ERD and ERS provide stable and consistent brain signals [31] and each can be used as an indicator of whether a user is identified or unidentified, resulting from individual differences in cognitive abilities [10]. Thus, it is reasonable to use CSP and ERD/S because they are extremely suitable for the MI data characteristics and have already been proven by previous studies [32], [33], [34]. Compared with other feature extraction methods, signals obtained from MI stimuli are easy to quantify and extremely helpful in improving user identification accuracy [35], [36].

In contrast, recent MI studies have tended to employ time series specialized feature extraction methods other than CSP and ERD/S for feature extractions, which have been known to enhance the time series characteristics of EEG using feature extraction methods, such as autoregressive (AR) [37], power spectral density (PSD) [38], and fast Fourier transform (FFT) [39]. Several studies have reported that AR and PSD were extracted features and then classified using SVM in the Graz dataset B, known as a dataset for motor imagery classification [37], [38]. Moreover, an identification accuracy of up to 95.89% was achieved between the imaginary motions of the tongue and the left little finger for an individual in the FFT-based MI data using SVM [40]. Lee et al. have reported that the time domain parameter (TDP) outperformed CSP in identification accuracy between hand grasping and wrist twisting, based on the MI tasks using SVM [41]; this phenomenon showed that TDP (93.6%) was more accurate than CSP (91.4%). Hence, these results contradict previous studies, which demonstrated that CSP and ERD/S were adequate as MI data-driven feature extraction methods [42], [43]. It is also still to be determined whether MI-driven data are suitable for user identification because previous studies have classified the imaginations that move the body.

In this study, we proposed an EEG-MI methodology, which had been optimized to identify each individual from groups using the 'Big Data of 2-classes MI' [44], which represents the most widely available large-scale dataset for MI analysis. To achieve this goal, we chose four feature extraction methods, which are frequently used in MI research: CSP, ERD/S, AR, and FFT. The extracted feature was classified between the user and non-users by SVM and GNB, and the calculated findings were compared against each other. Subsequently, a validation process using the half-total error rate (HTER) was also conducted to solve the imbalance problem caused by the difference in data sizes. As a result, we found that CSP provided the highest accuracy among all feature extraction methods and obtained accuracies of 98.97% and 97.47% in the SVM and GNB classifiers, respectively. Despite the high accuracy of CSP, it is also possible to detect between one user and the non-users using the 'dataset IVa' [45]—known as a small MI dataset—considering the sensitivity of CSP is weak in data sizes that are too small. In this process, we can estimate the minimum number of users required to ensure a reliable CSP performance, even with small sample

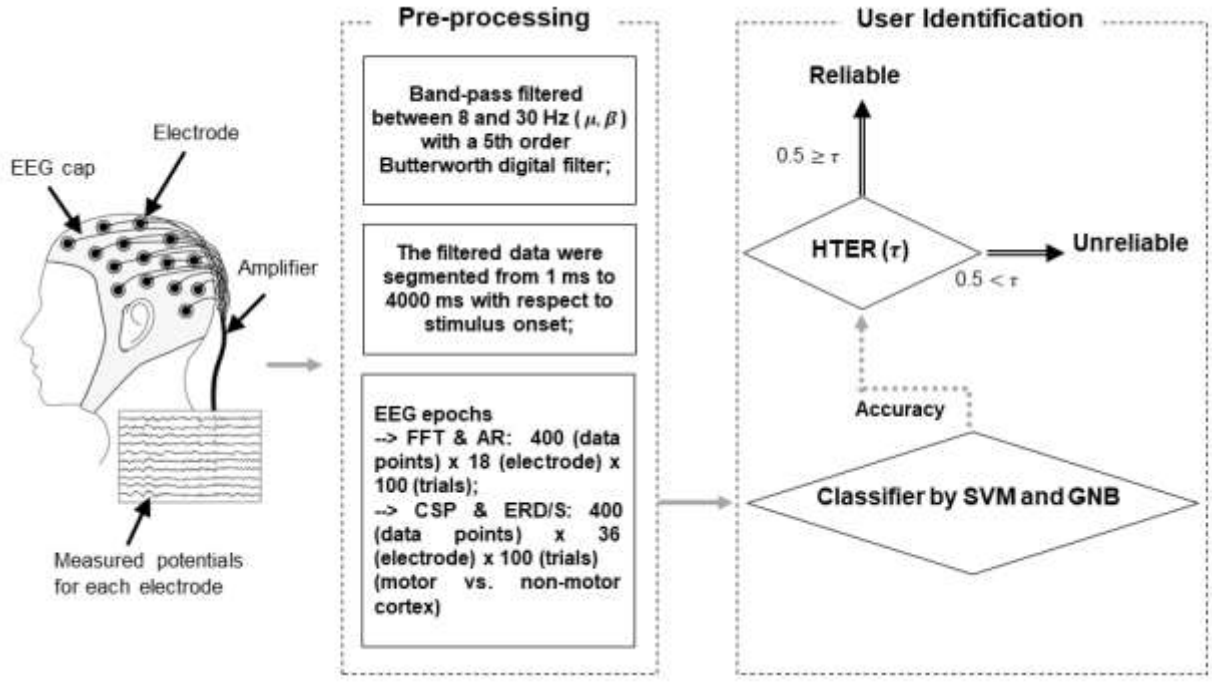



**FIGURE 1.** Overview of the process of user identity recognition based on EEG-MI signals. After performing the data pre-processing, including filtering, segmentation, and feature extraction, we divided the pre-processed data into training sets and test sets to avoid overfitting, which does not work well given new data. Then, the user identification methodology used SVM and GNB to calculate the accuracies for recognizing specific users among all users. The reliability of the calculated accuracy was determined by HTER. When this value is less than 0.5, the user identification accuracy is reliable. Therefore, we can obtain identification accuracy for all users through this process.

sizes. These results are confirmed by HTER. Therefore, this study is very helpful in implementing an EEG-based MI user identification, which improves the accuracy in recognizing users and guarantees user-aware performance.

## II. MATERIALS AND METHODS

To recognize a user by improving their identification accuracy, we proposed an EEG-MI methodology, which used the optimized feature extraction methods and classifiers based on EEG-MI datasets. Figure 1 shows the detailed procedures for the proposed methodology. During pre-processing, EEG signals were filtered, segmented, and the features extracted. Specifically, we used one of the large-scale MI datasets known as the 'Big Data of 2-classes MI' [44]. We extracted features using CSP, ERD/S, AR, and FFT; then, divided the extracted data into training sets and test sets to avoid overfitting [46]. Then, we detected the user by comparing identified accuracies using SVM and GNB. The two classifiers are known to operate in completely different modes, with SVM classifying the two classes by maximizing the margin, whereas GNB operates on probability, thus the performance differences between the two classifiers are considered clear. Finally, a two-independent sample t-test was performed to determine whether there is a statistically significant difference in the identified accuracies. To evaluate the reliability of the identified accuracy, HTER was calculated using the means of FAR and FRR, which are two errors in the biometric identification method.

### A. EEG DATASETS

According to G*Power 3.1.9.2, we need to collect at least 30 users. We assumed a moderate effect size ($\alpha = 0.05$, power=0.8, effect size=0.25) [47]. We adopted the publicly available EEG dataset used by Lee et al. [44], which is one of the largest datasets based on EEG-MI tasks. This dataset has been widely used in many previous studies [48], [49], [50] and has been well collected by proven system protocols [51]. The raw EEG signals were recorded from 54 healthy users (ages: 24–35; 29 males) by 62 channels with a sampling frequency of 1 kHz. To classify the EEG brain patterns between the motor and the non-motor regions, we selected 18 motor channels located in the parietal region and 18 non-motor channels located in the frontal and occipital regions. Figure 2 shows the selected channel locations and motor cortex positions. An overall experiment consists of a cue (3 s), a task imagining grasping either the right- or left-hands (4 s), and non-task (6 s) periods. A total of 100 trials were attempted. The acquired EEG signals are made up of various oscillatory activities, including the alpha and beta bands. These alpha and beta bands play an important role in determining brain states during MI tasks [52]. Thus, we performed band-pass filtering in the range of 8–30 Hz, which corresponded to the alpha and beta bands [53]. We segmented the task sections

ranging from 0 to 4 s for signal enhancement because of the repeated measurements. Therefore, we applied four feature extraction methods to the enhanced signals within the segmented sections.

### B. FOUR FEATURE EXTRACTION METHODS: CSP, ERD/S, FFT, AND AR

In this study, we used CSP, ERD/S, AR, and FFT to transform segmented data into informative features. The TDP method is excluded from this work because it is suitable for motor execution rather than motor imagination. Each method is described as follows.

#### 1) CSP

This method is a spatial filtering technique that takes into account spatial characteristics [54]. The CSP method is based on the simultaneous diagonalization of two covariance matrices. Simply, CSP maximizes the variance between classes. The spatially filtered signal $Z$ is given as:

$$Z = WE \tag{1}$$

where E is the N × T matrix representing raw EEG measurement data. Further, N is the number of channels, T is the number of measurement data per channel, and W is the CSP projection matrix. The rows of W are the stationary spatial filters and the columns of $W^{-1}$ are the common spatial patterns.

The spatially filtered signal Z given in Eq. (1) maximizes the differences in the variance of the two classes of EEG measurements. The spatially filtered signals are generally used as features for classification [55]. For each imagined movement direction, $m$ is the number of imagined movements as the row vectors of Z. Thus, the feature vectors are formed by $X_p$ ($p \in \{1, \ldots, 2m\}$) given in Eq. (2) as inputs to the classifier.

$$X_p = log(var(Z_p)/\sum_{p=1}^{2m} var(Z_p)) \tag{2}$$

where $X_p$ are the feature vectors that are used to calculate a linear classifier by taking log to have a normal distribution.

#### 2) ERD/S

The standard method for quantifying ERD/S was proposed by Pfurtscheller et al. [56]. This method shows the relative variability between the recorded EEG power during the non-task before the event occurred and the recorded EEG power during the event. To extract the features using the ERD/S method, the power spectrum of the non-task must have a clear peak. We performed the following procedures for the identification of the individuals. First, we obtained $x_f(t; i)$, the bandpass-filtered signal in *trial* $i \epsilon \{1, \ldots, N\}$, from $x(t; i)$, the EEG signals. The filtered signals were then squared to obtain the power signals, $S_f(t; i)$, as follows:

$$S_f(t; i) = x(t; i)^2(t; i) \tag{3}$$

The averaged power signal, $\bar{S}_f(t)$, obtained for the MI task for all trials, was calculated by:

$$\bar{S}_f(t) = \frac{1}{N}\sum_{i=1}^{N} S_f(t; i) \tag{4}$$

where $N$ is the total number of trials. Conversely, $R_f$, which was obtained from the non-task for a comparison relative to the MI task, was calculated by:

$$R_f = \frac{1}{N}\sum_{t \in t_b}^{k} S_f(t; i) \tag{5}$$

where $t_b \in \{t_1, \ldots, t_k\}$ is the non-motor imagery channel data. The relative power change, $C_{f(t)}$, can be calculated as follows:

$$C_{f(t)} = \frac{\bar{S}_f(t) - R_f}{R_f} \times 100(\%) \tag{6}$$

#### 3) AR

The AR model is a linear regression of the current observed data against the prior observed data. The AR model is calculated as:

$$x(t) = \sum_{i=1}^{p} a_i x(t - i) + \varepsilon_t \tag{7}$$

where $x(t)$ is the current observed data at the current discrete time $t$, which is dependent upon the data at the previous time steps. In other words, an AR model of order '$p$', indicates that the current observed values depend on the '$p$' past observed values. Further, $a_i$, are the model coefficients corresponding to the $i$ the order $x(t)$; $\varepsilon_t$ is the white noise at time $t$.

#### 4) FFT

FFT is a representative spectral analysis algorithm in the frequency domain, representing time-series signals as various power spectra of frequencies. FFT is often used as a benchmark to compare with other spectral analysis methods because of its simplicity and efficacy. FFT is calculated as:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}}, \quad k = 0, 1, \ldots N-1 \tag{8}$$

where $X_k$ is the FFT coefficients, $n$ is the total number of points in FFT and $N$ is the total number of input EEG data.

### C. USER IDENTIFICATION USING THE SVM AND GNB CLASSIFIERS

We calculated the accuracy of distinguishing one user from 53 non-users using two classifiers: SVM and GNB. SVM is widely used for pattern recognition, classification, and regression, as one of the supervised machine-learning algorithms. We used SVM with radial basis function (RBF) in this study. It is critical for SVM to find a hyperplane with the farthest distance between support vectors (i.e., each vector corresponds to two classes) in an infinite-dimensional space for data classification. The advantage of SVM is that it optimizes the weight with the lowest cost function [57]. This method is determined by whether the data for each class belongs to the user (Y=0) or belongs to the non-user (Y=1). Similarly, GNB can also be divided into user (Y=0) and

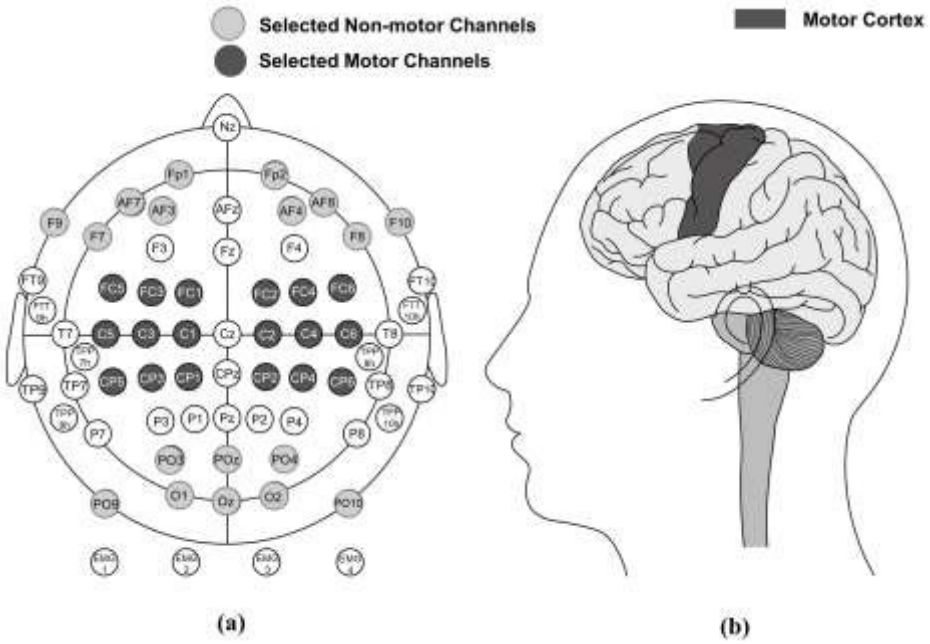



**FIGURE 2.** Location of the 62 electrodes and the predefined regions of interest (ROIs): (a) configuration of the EEG channels consisting of the motor area (dark area) and non-motor area (light-dark area) and (b) subject's head in the area of the motor cortex (dark area). The electrodes were placed on the scalp at AFz, according to the international 10-20 system.

non-user (Y=1), although it differs from SVM in that it is based on the Bayes hypothesis, having independence between predictors [58]. The GNB classifier recognizes the effect of the value of a previous probability of a predictor on a given target value [59]. The two models are widely used in user identification fields.

Through these two classifiers, we progressed through the process shown in Figure 3 to classify one user from the other 53 non-users. To avoid the overfitting or learning biases caused by the classifier's heavy dependency on the training data, the designed classifiers were tested using a 10-fold cross-validation. For cross-validation, we divided them into nine train sets and one test set. Each classifier model was tested 30 times and then the mean identification accuracy and the standard error of the mean (SEM) were calculated to estimate the variability of the accuracies. In both classifiers, however, the binary classification between one user and 53 non-users has the potential to cause an imbalance problem due to the difference in datasets between the two classes [60]. Therefore, we calculated HTER to determine whether the classifiers' results were affected by the imbalance problem in the following section.

### D. HTER AS AN INDICATOR FOR DETERMINING THE RELIABILITY OF CLASSIFIER'S RESULTS

There are two types of errors in a biometric system: false acceptance (FA), in which the user identification accepts a non-user, and false rejection (FR), in which the user identification rejects a genuine user. They are stated as a percentage of the false acceptance rate (FAR) and false rejection rate (FRR). They are averaged as HTER ($= (FAR + FRR)/2$) to objectively determine the reliability of the accuracy of the results calculated from the classifiers [25]. Ultimately, the lower the HTER value, the more reliable the user identification methodology based on the EEG biometrics is. Therefore, we considered the security criterion for user identification as the HTER value, which should be less than 0.5, otherwise, they were deemed a non-user.

### E. STATISTICAL ANALYSIS USING A TWO-INDEPENDENT SAMPLE T-TEST

A Statistical Package for the Social Sciences (SPSS), version 25.0 (IBM Corp., Chicago, Illinois, USA) was used for all statistical analyses. A Two-independent sample t-test is a method used to test whether the unknown population means of two groups are equal or not. The two-independent sample t-test was utilized to determine whether there were statistically significant differences in classified accuracy between a user and the non-users in the combinations based on four feature extraction methods and two classifiers. The values of $p < 0.05$ or $p < 0.001$ were considered statistically significant.

## III. EXPERIMENTAL RESULTS

### A. CLASSIFICATION ACCURACY BETWEEN ONE USER AND 53 NON-USERS WITH FOUR FEATURE EXTRACTION METHODS USING SVM AND GNB

We show the resulting ranking in order of accuracy based on four features and two classifier combinations. Figure 4 depicts the accuracies of user identification using SVM for

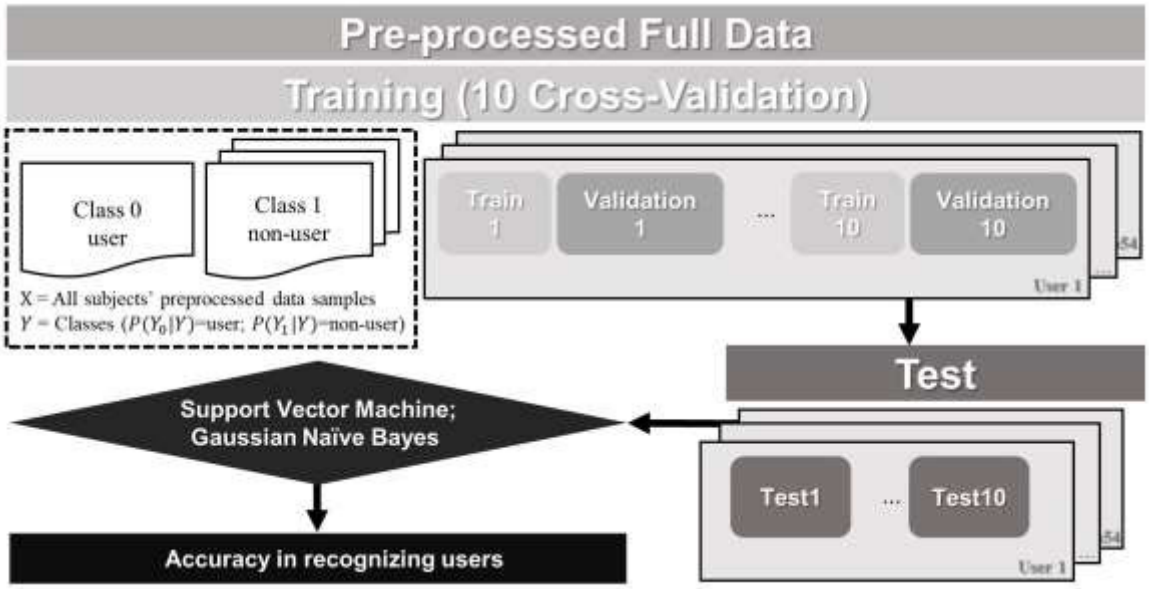



**FIGURE 3.** Schematic flow chart for user identification between a user and the non-users using the SVM and GNB classifiers. They were tested using 10-fold cross-validation to prevent overfitting or learning biases caused by the classifier's heavy dependency on training data. We divided the training data into nine train sets and one test set. Each classifier model was tested 30 times and then the accuracy for recognizing users and the standard error of the mean (SEM) were calculated to estimate the variability of the accuracies.

**TABLE 1.** HTERs based on the four feature extraction methods using SVM and GNB. HTERs are obtained to be less than 0.5, indicating that the calculated accuracies are reliable.

| Feature extraction method | HTER±SD using SVM | HTER+SD using GNB |
|---|---|---|
| CSP | $0.01 \pm 0.00$ | $0.02 \pm 0.01$ |
| ERD/S | $0.01 \pm 0.00$ | $0.03 \pm 0.03$ |
| FFT | $0.01 \pm 0.00$ | $0.18 \pm 0.12$ |
| AR | $0.02 \pm 0.00$ | $0.41 \pm 0.08$ |

each feature extraction method. The use of 'A' in the x-axis indicates the average accuracy for user identification between the one user and the 53 other non-users. The averaged identification accuracies appear in the order of CSP (98.97%), ERD/S (98.94%), AR (98.93%), and FFT (97.92%), highlighting that almost equal accuracies were observed for each extraction method. Likewise, Figure 5 shows the user identification accuracies using the GNB classifier. High accuracies were again obtained in the order of CSP (97.47%), ERD/S (94.58%), FFT (53.80%), and AR (50.24%). The accuracy difference between the highest accuracy (CSP) and the lowest accuracy (AR) was 47.23%. Thus, emphasizing that the CSP method is superior to other methods for both classifiers. This phenomenon can estimate that the GNB's features are influenced by one another and will affect the results organically.

Furthermore, Figure 6 shows the boxplots of the user identification accuracies of the 54 total users for the four feature methods using GNB and SVM. The CSP, using both GNB and SVM, shows the highest mean accuracy and the smallest variation in the accuracy distributions. In each boxplot, the red lines and red crosses represent the median and outliers of the distributions, respectively.

### B. STATISTICAL ANALYSIS USING A TWO-INDEPENDENT SAMPLE T-TEST

Figure 7 shows the user identification accuracies between one user and the non-users, and their HTERs in four feature extraction methods using SVM and GNB. Their differences were then statistically compared using a two-independent sample t-test. Before statistically comparing the differences in accuracy, we found that the two combinations among the four feature extraction methods clearly show significant differences in the user identification accuracies and their HTERs, as follows: FFT vs. CSP, FFT vs. ERD/S, AR vs. CSP, and AR vs. ERD/S. The asterisks indicate significant differences determined by a two-independent sample t-test ($***p < 0.001$; $*p < 0.05$) and the red error bars are SEMs to estimate the variability of the accuracies and HTER values.

Figure 7(a) illustrates the bar graphs showing the statistical differences in each combination based on the accuracy, as follows: FFT vs. CSP ($***p < 0.001$), FFT vs. ERD/S ($***p < 0.001$), AR vs. CSP ($*p < 0.05$), and AR vs. ERD/S ($***p < 0.001$), and result in significant differences in the accuracies for the combinations.

In all the feature combination cases, there are statistically significant HTER differences ($***p < 0.001$) between the two feature combinations, as shown in Figure 7(b). Thus, we discovered that there are significant differences in the

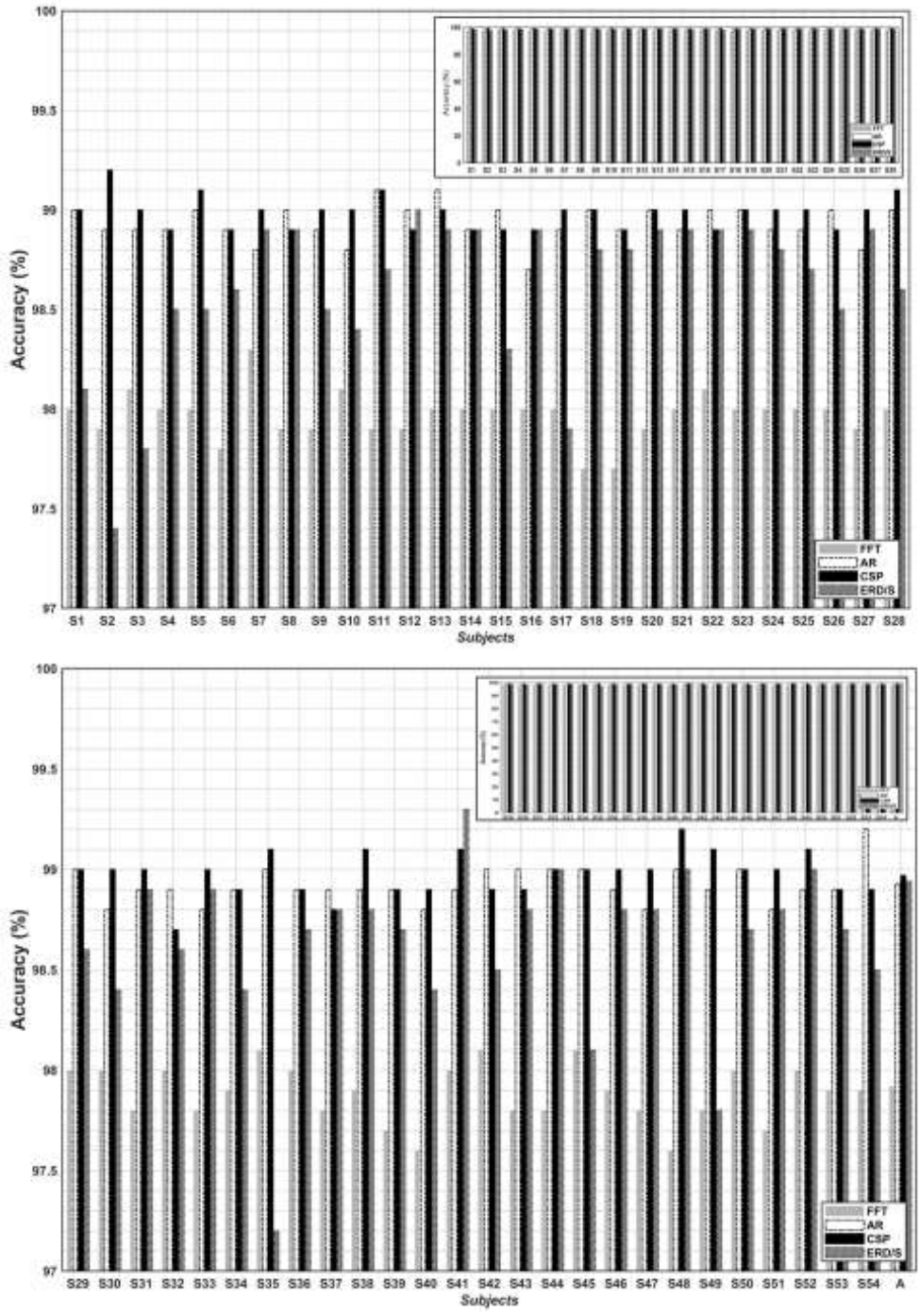



**FIGURE 4.** User identification accuracies of 54 users using SVM, in accordance with four feature extraction methods. 'A' in the x-axis indicates the average accuracy for recognizing a user across 54 users. The highest accuracy is derived in the order of CSP (98.97%), ERD/S (98.94%), AR (98.93%), and FFT (97.92%).

accuracies of identifying users and their HTER values in each feature extraction method.

### C. EEG IDENTIFICATION BASED ON A SMALL DATASET

We have achieved a sufficiently high user identification accuracy using the CSP-SVM method. However, when CSP is applied to small datasets, it tends to rely on sample-based covariance, which can lead to poorer accuracy [19]. Thus, the dataset size is known to be extremely sensitive to CSP.

To investigate whether the CSP method is actually affected by the size of the data, we tested a publicly known, small dataset—Dataset IVa [61]. We followed the same steps to calculate the identification accuracies for Dataset IVa using the CSP-SVM combination.

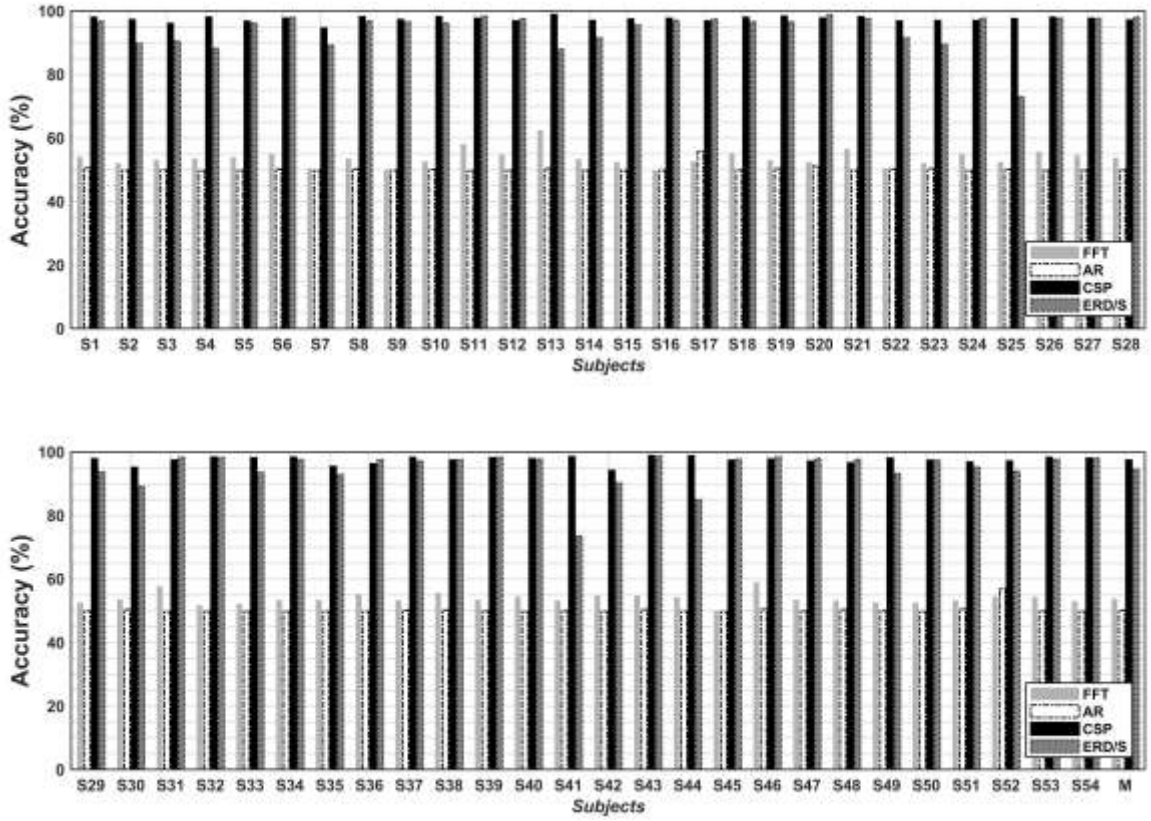



**FIGURE 5.** User identification accuracies of 54 users using GNB in accordance with four feature extraction methods. 'M' in the x-axis indicates the average accuracy for recognizing a user across 54 users. The highest accuracy is derived in the order of CSP (97.47%), ERD/S (94.58%), FFT (53.80%), and AR (50.24%).

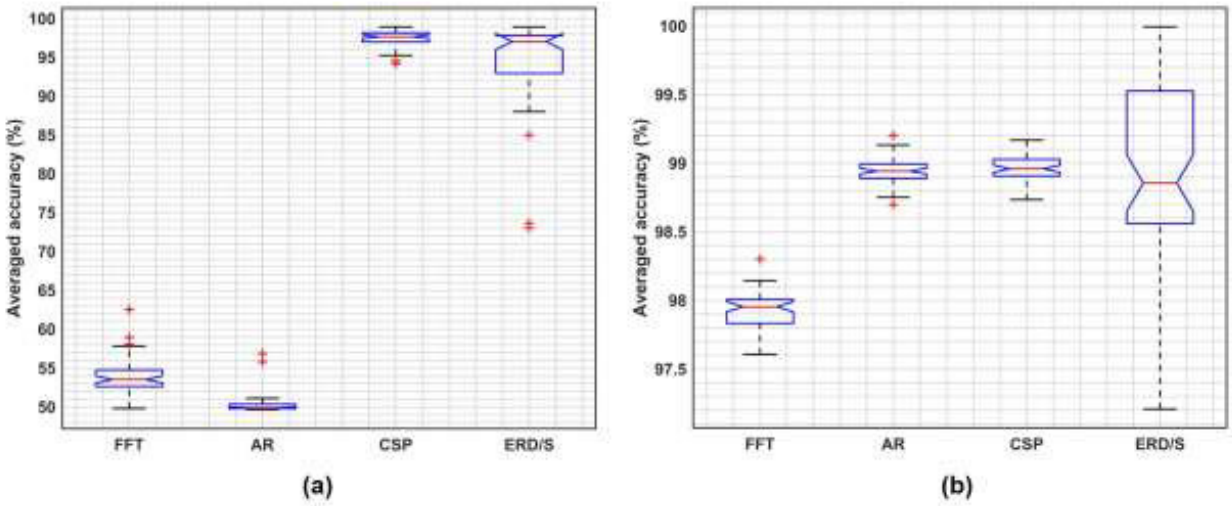



**FIGURE 6.** Boxplots of user identification accuracies of the 54 total users based on the four feature extraction methods using (a) GNB, and (b) SVM. CSP provides the highest accuracy and smallest variation in the accuracy distributions, allowing for consistent findings. The red lines and red crosses in each box represent the median and outliers of the distributions, respectively.

As a result, we achieved a user recognition accuracy of 56.2% in Dataset IVa. This result shows a poor performance compared to the dataset known as the 'Big Data of 2-classes MI' (98.9%). However, all the results were also guaranteed through the HTER and SD ($0.5 \pm 0.00$) in Dataset IVa, and the HTER and SD ($0.01 \pm 0.00$) for 'Big Data of 2-classes MI'.

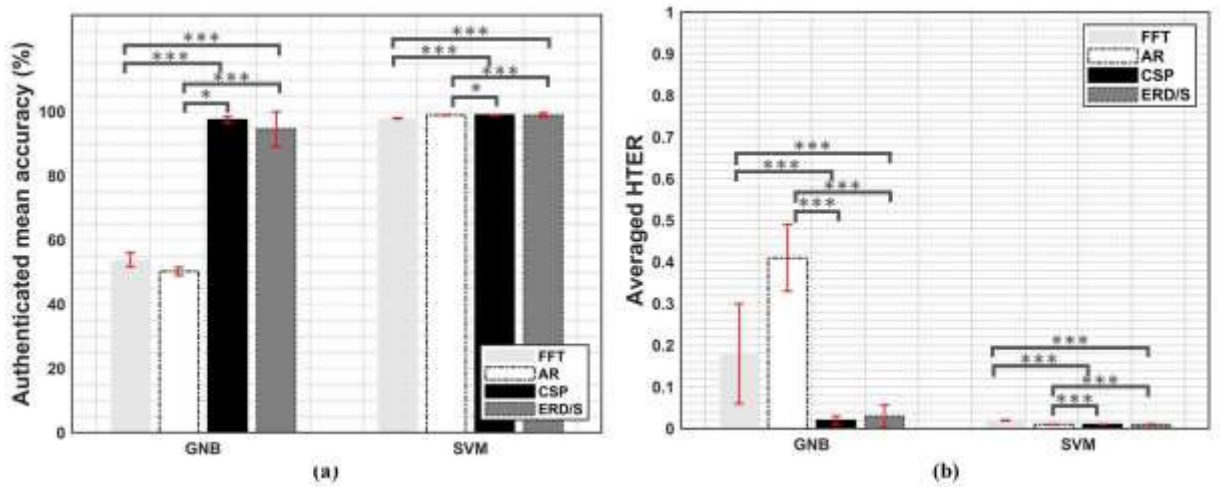



**FIGURE 7.** Bar graphs representing (a) user recognition accuracies and (b) their HTERs using a two-independent sample t-test, in accordance with the feature extraction methods using SVM and GNB. The red error bars are SEMs and the black horizontal thick lines show the significant levels between the two feature extraction methods, as indicated by the vertical short solid black lines at the end of the thick solid black horizontal lines. The asterisks show the significance levels ($*p < 0.05$; $***p < 0.001$). There are significant statistical differences in user identification accuracies and their HTERs. There are significant differences in the authenticated mean accuracies and the averaged HTERs (FFT vs. CSP, FFT vs. ERD/S, AR vs. CSP, AR vs. ERD/S).

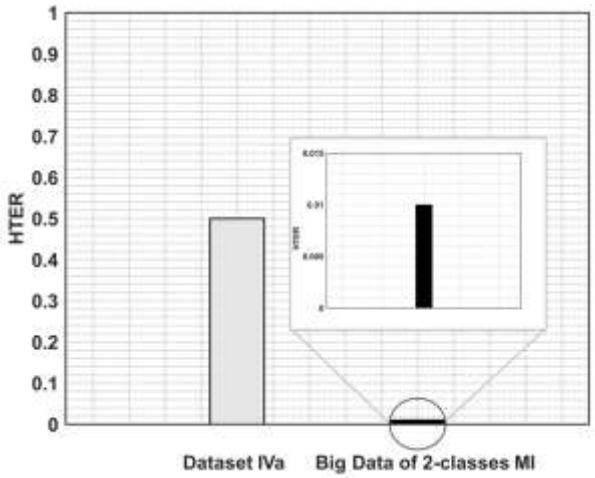



**FIGURE 8.** Comparison of HTERs calculated from small-scale user datasets (Dataset IVa) and large-scale user datasets (Big Data of 2-classes MI) using the CSP-SVM approach. The user identification accuracies of both datasets are reliable by HTERs. However, higher reliability is guaranteed in the large-scale user dataset due to the low HTER of the large-scale user datasets ($0.01 \pm 0.00$) over the small user datasets ($0.5 \pm 0.00$). Although the user identification accuracy in Dataset IVa is distinctly inferior, we ensured the results of the user identification accuracy.

These are illustrated in Figure 8. Moreover, it means that the accuracies of both datasets are guaranteed to be the HTER less than 0.5.

### D. RELIABILITY OF USER IDENTIFICATION ACCURACIES GUARANTEED BY HTER

To assess the reliability of the identified accuracies, we used the HTERs as the threshold ($\tau$) for security settings. The HTERs range from zero to one and numbers closer to zero represents higher reliability. Table 1 shows the averaged HTERs and their standard deviations (SD) using SVM and GNB. SVM presents an almost equal HTER and SD. However, GNB shows the low-reliability order of CSP ($0.02 \pm 0.01$), ERD/S ($0.03 \pm 0.03$), FFT ($0.18 \pm 0.12$), and AR ($0.41 \pm 0.08$). These results guarantee sufficient reliability of the calculated accuracies because their HTER values are below 0.5.

### E. PROPOSED METHODOLOGY FOR ESTIMATING THE NUMBER OF USERS BASED ON EEG-MI SIGNALS

In the dataset of 'Big Data of 2-classes MI' [44], we were successful in achieving a high accuracy in the classification between the one user and the non-users. Then, we additionally tested a small dataset, Dataset IVa, with the same processes and obtained a lower accuracy [61]. Dataset IVa, a small-sample dataset consisting of five brain signals, is likely to degrade the performance [62], [63], [64]. Furthermore, the CSP, which depends on sample-based covariance, finds it difficult to yield highly recognized accuracy on small-scale datasets [19]. Therefore, we proposed a method for estimating the number of recognizable users (i.e., data scales) and ensuring the results obtained from biometric systems for small datasets. For this purpose, we defined the threshold for security settings as HTER ($\tau$) and selected its value as between 0 and 0.5. Then, we calculated the mean of the HTERs ($\bar{X}$), borrowing the concept of basic statistics [65], as follows:

$$\bar{X} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) \tag{9}$$

where $n$ is the total number of users and $x$ refers to the HTER value calculated from each user. We randomly selected two

HTERs out of all the users' HTERs and calculated $\bar{X}$ with Eq. (9). This process was repeated until $\bar{X}$ reached $\tau$.

This method was then applied to Dataset IVa. It is possible to predict up to five users when a threshold of 0.5 ($\tau = 0.5$) is established for the security settings. In other words, this estimated result ensures reliability, while also satisfying the HTER value. Hence, this study estimated the number of users regardless of the small sample size, while also ensuring the reliability of the user identification method.

## IV. DISCUSSION

### A. EEG-MI METHODOLOGY WITH A HIGHLY RECOGNIZABLE ACCURACY

Personal identification models have recently been developed using time series data-based feature extraction methods in security fields, such as AR and PSD, rather than CSP [66]. However, our study demonstrated that the EEG-MI methodology is suitable for biometric systems owing to its high user identification accuracy. This methodology was utilized with the CSP feature extraction method and SVM classifier based on the EEG-MI datasets, as shown in Figures 4–7. Similarly, T. N. Alotaiby et al. showed that the CSP–SVM combination provides a highly recognizable user accuracy (95.15%) [67], although our results exhibited a 3.82% higher accuracy than those in their study. Furthermore, our results can reliably ensure personal identification, exhibiting a uniformly high accuracy across all 54 subjects (Figures 4–7).

### B. DIFFICULTY CLASSIFYING THE ONE USER FROM THE NON-USERS

We used the CSP–SVM methodology in the security field, which signifies one of the ways to identify a user from multiple users based on EEG-MI signals. Studies relating to our proposed model are not common because most studies focus on identifying other biological signals, such as the iris and fingerprints [68], [69], rather than brain waves. Many researchers also focused on classifying the brain signals recorded between cognitive tasks and non-tasks as performed by a user [25], [70]. Moreover, previous studies have provided relatively low accuracy: Andreas M. Ray et al. previously reported that the use of 27 healthy subjects achieved a mean classification accuracy of 75.30% [71]. A recent study obtained an average classification accuracy of 71.20% from a total of 18 participants [72]. However, they had fewer subjects than our case and their average accuracies were 27.77% and 26.27% lower than the average accuracies obtained from the SVM and GNB models in this study, respectively (Figure 4–5). Therefore, our study demonstrated, for the first time, how to identify only one user from multiple non-users using the CSP–SVM combination, based on EEG-MI.

### C. PROPOSED METHODOLOGY FOR ESTIMATING THE NUMBER OF USERS

To compensate for the shortcomings of CSP, which is less accurate when using small datasets, we estimated the minimum number of users required to ensure the reliability of the classification results using Eq. (9). We confirmed that a data size of five people is a viable data size for ensuring the appropriate reliability by using the Dataset IVa—a small dataset. Figure 8 demonstrates the reliable HTER results after using the CSP–SVM combination. It is appropriately reasonable to estimate the minimum number of users since the data size is affected by the number of users, in terms of information security. Many studies have already reported the poor accuracy of the results when the data size is too small [62], [63], [64]. It is also important in the field of security to prevent the exposure of the user's personal information [73], alongside improving on the accuracy of their identification [74]. Therefore, the number of users is an important factor that can affect the security system; this study was able to predict the required data size by demonstrating the reliability of HTERs. Hence, our proposed method is simple and is expected to make a significant contribution to the security field.

## V. CONCLUSION

We proposed the EEG-MI methodology for user identification to improve the accuracy of identifying one user from other non-users and ultimately guarantee their reliability. We compared four feature extraction methods along with optimized classifiers. The CSP–SVM method, in particular, provided the best identification performance with an accuracy of 98.97%, compared to any other combination method. Moreover, the GNB classifier produced the highest identification difference between CSP, with the highest accuracy (97.47%), and AR, with the lowest accuracy (50.24%), which resulted in an accuracy difference of 47.23%. Despite the high performances of using CSP, it is very sensitive to data size. Thus, we can estimate the dataset size required to ensure CSP performance, even with small data scales. Lastly, all user recognition results were evaluated by HTER to determine whether they were affected by the imbalance problem in binary classifications; all our results are reliable. This study contributed to developing an optimized user identification method for the EEG-MI methodology by significantly improving user identification accuracy.

## ETHICAL STATEMENT

No ethical statement to be declared

## AUTHORS CONTRIBUTION

Conceptualization: Sujin Bak; methodology: Sujin Bak; software: Sujin Bak; validation: Sujin Bak; formal analysis: Sujin Bak; investigation: Sujin Bak; resources: Sujin Bak; data curation: Sujin Bak; writing—existing draft preparation: Sujin Bak and Jichai Jeong; writing—review and editing: Sujin Bak and Jichai Jeong; visualization: Sujin Bak; supervision: Jichai Jeong; project administration: Jichai Jeong; funding acquisition: Sujin Bak and Jichai Jeong. All authors have read and agreed to the published version of the manuscript.

## ACKNOWLEDGMENT

The authors would like to express their gratitude to Essayreview (www.essayreview.co.kr) for performing the English language editing.

## REFERENCES

[1] C. Murukesh, K. Thanushkodi, and P. Padmanabhan, "Secured authentication through integration of gait and footprint for human identification," *J. Electr. Eng. Technol.*, vol. 9, no. 6, pp. 2118–2125, 2014.

[2] T. Ring, "Spoofing: Are the hackers beating biometrics?" *Biometric Technol. Today*, vol. 2015, no. 7, pp. 5–9, Jul. 2015.

[3] A. Roy, N. Memon, J. Togelius, and A. Ross, "Evolutionary methods for generating synthetic masterprint templates: Dictionary attack in fingerprint recognition," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 39–46, doi: 10.1109/ICB2018.2018.00017.

[4] Z. Wei, X. Qiu, Z. Sun, and T. Tan, "Counterfeit iris detection based on texture analysis," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[5] V. Ruiz-Albacete, P. Tome-Gonzalez, F. Alonso-Fernandez, J. Galbally, J. Fierrez, and J. Ortega-Garcia, "Direct attacks using fake images in iris verification," in *Proc. Eur. Workshop Biometrics Identity Manage.* Cham, Switzerland: Springer, 2008, pp. 181–190.

[6] M. Abo-Zahhad, S. M. Ahmed, and S. N. Abbas, "A new multi-level approach to EEG based human authentication using eye blinking," *Pattern Recognit. Lett.*, vol. 82, pp. 216–225, Oct. 2016.

[7] Z. Mu, J. Hu, and J. Min, "EEG-based person authentication using a fuzzy entropy-related approach with two electrodes," *Entropy*, vol. 18, no. 12, p. 432, 2016.

[8] Y. Zeng, Q. Wu, K. Yang, L. Tong, B. Yan, J. Shu, and D. Yao, "EEG-based identity authentication framework using face rapid serial visual presentation with optimized channels," *Sensors*, vol. 19, no. 1, p. 6, Dec. 2018.

[9] D. J. A. Smit, D. Posthuma, D. I. Boomsma, and E. J. C. Geus, "Heritability of background EEG across the power spectrum," *Psychophysiology*, vol. 42, no. 6, pp. 691–697, Nov. 2005.

[10] Z. A. A. Alyasseri, A. T. Khader, M. A. Al-Betar, J. P. Papa, and O. A. Alomari, "EEG feature extraction for person identification using wavelet decomposition and multi-objective flower pollination algorithm," *IEEE Access*, vol. 6, pp. 76007–76024, 2018.

[11] Z. A. A. Alyasseri, O. A. Alomari, S. N. Makhadmeh, S. Mirjalili, M. A. Al-Betar, S. Abdullah, N. S. Ali, J. P. Papa, D. Rodrigues, and A. K. Abasi, "EEG channel selection for person identification using binary grey wolf optimizer," *IEEE Access*, vol. 10, pp. 10500–10513, 2022.

[12] M. Ienca and P. Haselager, "Hacking the brain: Brain–computer interfacing technology and the ethics of neurosecurity," *Ethics Inf. Technol.*, vol. 18, no. 2, pp. 117–129, Jun. 2016.

[13] J. Sohankar, K. Sadeghi, A. Banerjee, and S. K. Gupta, "E-BIAS: A pervasive EEG-based identification and authentication system," in *Proc. 11th ACM Symp. QoS Secur. Wireless Mobile Netw.*, 2015, pp. 165–172.

[14] W. Huang, X. Chen, R. Jin, and N. Lau, "Detecting cognitive hacking in visual inspection with physiological measurements," *Appl. Ergonom.*, vol. 84, Apr. 2020, Art. no. 103022.

[15] C. T. Lin, C.-H. Chuang, Z. Cao, A. K. Sing, C.-S. Hung, Z. Cao, A. K. Sing, C.-S. Hung, Y.-H. Yu, M. Nascimben, Y.-T. Liu, J.-T. King, T.-P. Su, and S.-J. Wang, "Forehead EEG in support of future feasible personal healthcare solutions: Sleep management, headache prevention, and depression treatment," *IEEE Access*, vol. 5, pp. 10612–10621, 2017.

[16] A. R. Elshenaway and S. K. Guirguis, "Adaptive thresholds of EEG brain signals for IoT devices authentication," *IEEE Access*, vol. 9, pp. 100294–100307, 2021.

[17] R. Shahzadi, S. M. Anwar, F. Qamar, M. Ali, J. J. P. C. Rodrigues, and M. Alnowami, "Secure EEG signal transmission for remote health monitoring using optical chaos," *IEEE Access*, vol. 7, pp. 57769–57778, 2019.

[18] S.-L. Sun, J.-H. Xu, L.-Y. Yu, Y.-G. Chen, and A.-L. Fang, "Mixtures of common spatial patterns for feature extraction of EEG signals," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Jul. 2008, pp. 2923–2926.

[19] H. Lu, H.-W. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2936–2946, Dec. 2010.

[20] E. A. Mousavi, J. J. Maller, P. B. Fitzgerald, and B. J. Lithgow, "Wavelet common spatial pattern in asynchronous offline brain computer interfaces," *Biomed. Signal Process. Control*, vol. 6, no. 2, pp. 121–128, Apr. 2011.

[21] K. Kalaivani and R. Sivakumar, "A novel fuzzy based bio-key management scheme for medical data security," *J. Electr. Eng. Technol.*, vol. 11, no. 5, pp. 1509–1518, Sep. 2016.

[22] Z. Mu, J. Yin, and J. Hu, "Application of a brain–computer interface for person authentication using EEG responses to photo stimuli," *J. Integrative Neurosci.*, vol. 17, no. 1, pp. 113–124, Feb. 2018.

[23] J.-F. Hu, "New biometric approach based on motor imagery EEG signals," in *Proc. Int. Conf. Future Biomed. Inf. Eng. (FBIE)*, Dec. 2009, pp. 94–97.

[24] R. Das, E. Maiorana, and P. Campisi, "Motor imagery for EEG biometrics using convolutional neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2062–2066.

[25] S. Marcel and J. D. R. Millan, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 743–752, Apr. 2007.

[26] K.-B. Lee, K. K. Kim, J. Song, J. Ryu, Y. Kim, and C. Park, "Estimation of brain connectivity during motor imagery tasks using noise-assisted multivariate empirical mode decomposition," *J. Electr. Eng. Technol.*, vol. 11, no. 6, pp. 1812–1824, Nov. 2016.

[27] C. Neuper, R. Scherer, M. Reiner, and G. Pfurtscheller, "Imagery of motor actions: Differential effects of kinesthetic and visual–motor mode of imagery in single-trial EEG," *Cognit. Brain Res.*, vol. 25, no. 3, pp. 668–677, Dec. 2005.

[28] G. Pfurtscheller and C. Neuper, "Future prospects of ERD/ERS in the context of brain–computer interface (BCI) developments," in *Progress in Brain Research*, vol. 159, C. Neuper and W. Klimesch Eds. Amsterdam, The Netherlands: Elsevier, 2006, pp. 433–437.

[29] G. Pfurtscheller, M. Woertz, G. Krausz, and C. Neuper, "Distinction of different fingers by the frequency of stimulus induced beta oscillations in the human EEG," *Neurosci. Lett.*, vol. 307, no. 1, pp. 49–52, Jul. 2001.

[30] M. Taniguchi, A. Kato, N. Fujita, M. Hirata, H. Tanaka, T. Kihara, H. Ninomiya, N. Hirabuki, H. Nakamura, S. E. Robinson, D. Cheyne, and T. Yoshimine, "Movement-related desynchronization of the cerebral cortex studied with spatially filtered magnetoencephalography," *NeuroImage*, vol. 12, no. 3, pp. 298–306, Sep. 2000.

[31] G. Pfurtscheller and F. L. Da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.

[32] M.-A. Li, J.-F. Han, and L.-J. Duan, "A novel MI-EEG imaging with the location information of electrodes," *IEEE Access*, vol. 8, pp. 3197–3211, 2020.

[33] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, "A CSPAM-BA-SVM approach for motor imagery BCI system," *IEEE Access*, vol. 6, pp. 49192–49208, 2018.

[34] Y. Park and W. Chung, "Optimal channel selection using correlation coefficient for CSP based EEG classification," *IEEE Access*, vol. 8, pp. 111514–111521, 2020.

[35] L. Liu, "Recognition and analysis of motor imagery EEG signal based on improved BP neural network," *IEEE Access*, vol. 7, pp. 47794–47803, 2019.

[36] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification," *IEEE Access*, vol. 7, pp. 18940–18950, 2019.

[37] T. Pham, W. Ma, D. Tran, P. Nguyen, and D. Phung, "A study on the feasibility of using EEG signals for authentication purpose," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2013, pp. 562–569.

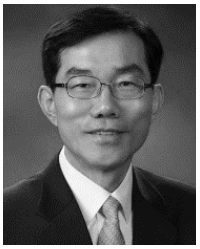
[38] P. Nguyen, D. Tran, X. Huang, and W. Ma, "Motor imagery EEG-based person verification," in *Proc. Int. Work-Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2013, pp. 430–438.

[39] D. Lee, H.-J. Lee, and S.-G. Lee, "Motor imagery EEG classification method using EMD and FFT," *J. KIISE*, vol. 41, no. 12, pp. 1050–1057, Dec. 2014.

[40] M. N. Islam, N. Sulaiman, M. Rashid, B. S. Bari, M. J. Hasan, M. Mustafa, and M. S. Jadin, "Empirical mode decomposition coupled with fast Fourier transform based feature extraction method for motor imagery tasks classification," in *Proc. IEEE 10th Int. Conf. Syst. Eng. Technol. (ICSET)*, Nov. 2020, pp. 256–261.

[41] S.-B. Lee, H.-J. Kim, H. Kim, J.-H. Jeong, S.-W. Lee, and D.-J. Kim, "Comparative analysis of features extracted from EEG spatial, spectral and temporal domains for binary and multiclass motor imagery classification," *Inf. Sci.*, vol. 502, pp. 190–200, Oct. 2019.
[42] Z.-C. Tang, C. Li, J.-F. Wu, P.-C. Liu, and S.-W. Cheng, "Classification of EEG-based single-trial motor imagery tasks using a B-CSP method for BCI," *Frontiers Inf. Technol. Electron. Eng.*, vol. 20, no. 8, pp. 1087–1098, Aug. 2019.
[43] Y. Park and W. Chung, "Frequency-optimized local region common spatial pattern approach for motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1378–1388, Jul. 2019.
[44] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, and S.-W. Lee, "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, May 2019, Art. no. giz002.
[45] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.
[46] Y. Zhang, J. Li, Y. Guo, C. Xu, J. Bao, and Y. Song, "Vehicle driving behavior recognition based on multi-view convolutional neural network with joint data augmentation," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4223–4234, May 2019.
[47] S. Bak, Y. Jeong, M. Yeu, and J. Jeong, "Brain–computer interface to predict impulse buying behavior using functional near-infrared spectroscopy," *Sci. Rep.*, vol. 12, no. 1, p. 18024, Oct. 2022.
[48] H. Altaheri, G. Muhammad, M. Alsulaiman, S. U. Amin, G. A. Altuwaijri, W. Abdul, M. A. Bencherif, and M. Faisal, "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," *Neural Comput. Appl.*, vol. 2021, pp. 1–42, Aug. 2021.
[49] Y.-E. Lee, G.-H. Shin, M. Lee, and S.-W. Lee, "Mobile BCI dataset of scalp- and ear-EEGs with ERP and SSVEP paradigms while standing, walking, and running," *Sci. Data*, vol. 8, no. 1, pp. 1–12, Dec. 2021.
[50] J.-H. Cho, J.-H. Jeong, and S.-W. Lee, "NeuroGrasp: Real-time EEG classification of high-level motor imagery tasks using a dual-stage deep learning framework," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13279–13292, Dec. 2022.
[51] G. Purtscheller and C. Neuper, "Motor imagery and direct brain–computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.
[52] H. Yuan, T. Liu, R. Szarkowski, C. Rios, J. Ashe, and B. He, "Negative covariation between task-related responses in alpha/beta-band activity and BOLD in human sensorimotor cortex: An EEG and fMRI study of motor imagery and movements," *NeuroImage*, vol. 49, no. 3, pp. 2596–2606, 2010.
[53] Y. Jeon, C. S. Nam, Y.-J. Kim, and M. C. Whang, "Event-related (De)synchronization (ERD/ERS) during motor imagery tasks: Implications for brain–computer interfaces," *Int. J. Ind. Ergonom.*, vol. 41, no. 5, pp. 428–436, Sep. 2011.
[54] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, pp. 787–798, May 1999.
[55] D. R. de Moraes Piazentin and J. L. G. Rosa, "Motor imagery classification for brain–computer interfaces through a chaotic neural network," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 4103–4108.
[56] H. Sun, L. Bi, and B. Chen, "Partitioned common spatial pattern method for single trial EEG signal classification in brain–computer interface system," *Automatika*, vol. 57, no. 1, pp. 66–75, Jan. 2016.
[57] M. A. Lopez-Gordo, D. Sanchez-Morillo, and F. P. Valle, "Dry EEG electrodes," *Sensors*, vol. 14, no. 7, pp. 12847–12870, 2014.
[58] M. Thyer, B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan, "Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis," *Water Resour. Res.*, vol. 45, no. 12, pp. 1–10, Dec. 2009.
[59] E. Frank, L. Trigg, G. Holmes, and I. H. Witten, "Naive Bayes for regression," *Mach. Learn.*, vol. 41, no. 1, pp. 5–25, 2000.
[60] A. Menon, H. Narasimhan, S. Agarwal, and S. Chawla, "On the statistical consistency of algorithms for binary classification under class imbalance," in *Proc. 30th Int. Conf. Mach. Learn.*, in Proceedings of Machine Learning Research, 2013, pp. 1–10. [Online]. Available: https://proceedings.mlr.press/v28/menon13a.html
[61] A. Singh, S. Lal, and H. W. Guesgen, "Small sample motor imagery classification using regularized Riemannian features," *IEEE Access*, vol. 7, pp. 46858–46869, 2019.
[62] Š. Raudys and A. K. Jain, "Small sample size problems in designing artificial neural networks," in *Machine Intelligence and Pattern Recognition*, vol. 11. Amsterdam, The Netherlands: Elsevier, 1991, pp. 33–50.
[63] E. Combrisson and K. Jerbi, "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy," *J. Neurosci. Methods*, vol. 250, pp. 126–136, Jul. 2015.
[64] L. Kanal and B. Chandrasekaran, "On dimensionality and sample size in statistical pattern classification," *Pattern Recognit.*, vol. 3, no. 3, pp. 225–234, Oct. 1971.
[65] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*. Boston, MA, USA: Cengage Learning, 2011.
[66] C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein, "Low-cost electroencephalogram (EEG) based authentication," in *Proc. 5th Int. IEEE/EMBS Conf. Neural Eng.*, Apr. 2011, pp. 442–445.
[67] T. N. Alotaiby, S. A. Alshebeili, L. M. Aljafar, and W. M. Alsabhan, "ECG-based subject identification using common spatial pattern and SVM," *J. Sensors*, vol. 2019, pp. 1–9, Mar. 2019.
[68] Y. Moolla, A. De Kock, G. Mabuza-Hocquet, C. S. Ntshangase, N. Nelufule, and P. Khanyile, "Biometric recognition of infants using fingerprint, iris, and ear biometrics," *IEEE Access*, vol. 9, pp. 38269–38286, 2021.
[69] A. Boyd, S. Yadav, T. Swearingen, A. Kuehlkamp, M. Trokielewicz, E. Benjamin, P. Maciejewicz, D. Chute, A. Ross, P. Flynn, K. Bowyer, and A. Czajka, "Post-mortem iris recognition—A survey and assessment of the state of the art," *IEEE Access*, vol. 8, pp. 136570–136593, 2020.
[70] H. Y. Yap, Y.-H. Choo, Z. I. M. Yusoh, and W. H. Khoh, "Person authentication based on eye-closed and visual stimulation using EEG signals," *Brain Informat.*, vol. 8, no. 1, pp. 1–13, Dec. 2021.
[71] A. M. Ray, R. Sitaram, M. Rana, E. Pasqualotto, K. Buyukturkoglu, C. Guan, and K.-K. Ang, "A subject-independent pattern-based brain–computer interface," *Frontiers Behav. Neurosci.*, vol. 9, p. 269, Oct. 2015.
[72] J. Kwon and C.-H. Im, "Subject-independent functional near-infrared spectroscopy-based brain–computer interfaces based on convolutional neural networks," *Frontiers Hum. Neurosci.*, vol. 15, Mar. 2021, Art. no. 646915.
[73] E. Aghasian, S. Garg, L. Gao, S. Yu, and J. Montgomery, "Scoring users' privacy disclosure across multiple online social networks," *IEEE Access*, vol. 5, pp. 13118–13130, 2017.
[74] M. Hammad, Y. Liu, and K. Wang, "Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint," *IEEE Access*, vol. 7, pp. 26527–26542, 2018.

**SUJIN BAK** received the B.S. degree in computer engineering from Jeonbuk National University and the M.S. and Ph.D. degrees in brain and cognitive engineering from Korea University, South Korea. She is currently with the Advanced Institute of Convergence Technology, Gyeonggi-do. Her main research interests include bio-signal processing and data analysis by artificial intelligence (AI) based on electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS).



**JICHAI JEONG** (Senior Member, IEEE) received the B.S. degree from Korea University, Seoul, South Korea, in 1980, the M.S. degree from KAIST, Daejeon, South Korea, in 1982, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 1988. From 1982 to 1985, he was with the Korea Institute of Science and Technology. From 1988 to 1993, he was a member of the Technical Staff with AT&T Bell Laboratories, Murray Hill, NJ, USA, where he worked on optoelectronic integrated circuits and semiconductor lasers for optical communications. From 1995 to 2007, he was a Faculty Member with the Department of Radio Engineering, Korea University. Since 2008, he has been with the Department of Brain and Cognitive Engineering, Korea University, where he is currently a Professor. His current research interests include functional infrared biophotonic spectroscopy, optical coherence tomography, and adaptive optical fiber sensing.